

AD-A122 650

READABILITY FORMULAS: THEIR APPLICATION IN THE ARMED  
FORCES(U) NAVY PERSONNEL RESEARCH AND DEVELOPMENT  
CENTER SAN DIEGO CA T M DUFFY NOV 82 NPRDC-RR-83-8

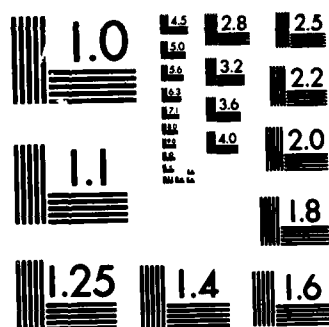
1/1

UNCLASSIFIED

F/G 5/2

NL


END  
FORMED  
DTIC



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

12

NPRDC SR 83-8

NOVEMBER 1982

AD A 122650

**READABILITY FORMULAS: THEIR APPLICATION  
IN THE ARMED FORCES**



**DTIC**  
**ELECTE**  
**DEC 22 1982**  
**S B D**

DTIC FILE COPY

**NAVY PERSONNEL RESEARCH  
AND  
DEVELOPMENT CENTER  
San Diego, California 92152**

**DISTRIBUTION STATEMENT A**

Approved for public release;  
Distribution Unlimited



22 011

**READABILITY FORMULAS: THEIR APPLICATION IN  
THE ARMED FORCES**

**Thomas M. Duffy**

**Reviewed by  
E. G. Aiken**

**Released by  
James F. Kelly, Jr.  
Commanding Officer**

**Navy Personnel Research and Development Center  
San Diego, California 92152**



## FOREWORD

This research was performed under exploratory development task area ZF63-522-011 (Assessment and Enhancement of Prerequisite Skills), work unit number ZF63-522-001-011-03.01 (Language Skills: Assessment and Enhancement). The report describes the limitations of readability formulas and proposes alternative methods for determining the comprehension requirements of Navy text and for ensuring that writers attend to comprehension requirements in producing new text. The issues and conclusions should be of particular interest to anyone involved in the procurement or production of training, job, or general information texts or manuals.

JAMES F. KELLY, JR.  
Commanding Officer

JAMES W. TWEEDDALE  
Technical Director

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	



## **SUMMARY**

### **Problem**

The armed forces have turned increasingly to the use of readability formulas to predict the ease with which users will be able to understand the text in technical manuals (TMs). Readability formulas are inexpensive, objective, and easy to use. They are, however, only proxies for the direct measurement of comprehension and they have the attendant weaknesses of proxies. The validity of reading grade level (RGL) scores obtained by these formulas is questionable, as is the use of these scores in determining the usability of military TMs.

### **Purpose**

The purpose of this work was to review and evaluate the use of readability formulas by the military.

### **Using Readability Formulas to Predict Literacy Gaps**

The primary use of readability formulas has been to predict the reading skill that will be required to use an existing manual. This has usually been done to predict whether or not there will be a "literacy gap"; that is, to determine whether the manual is written at too high a level for the intended users. This application requires that the formula identify a specific reading skill requirement that can be compared to the reading skill level of the users.

This application is clearly invalid and could seriously mislead writers as to the acceptability of their material, because the comprehension tasks and reading context involved in developing the formulas are so very different from the comprehension tasks and contexts for which predictions are being made.

If readability formulas are to be used to predict reading skill requirements, then new formulas based on realistic reading situations will have to be developed. Separate formulas may well be required for classroom training, self-paced training, job use, etc.

### **Using Readability Scores to Guide Production of Text**

Readability scores derived from various formulas have been used both as criteria for guiding the production of TMs and as binding specifications. In practice, readability scores will be ineffective for these purposes simply because, under the time and financial pressures involved in TM production, writers are forced to write to the formula (e.g., to select a word primarily because it is short, not because it is best for understanding).

The use of readability scores as guidelines rather than as specifications should reduce the tendency to write to formula, while still encouraging a focus on readable writing practices (e.g., simplification of words and sentences).

A review of the literature indicates that, when writing is revised in accordance with readable writing guidelines, practical effects on the level of performance in actual comprehension tests are rare. Even when use of the guidelines has had an effect, the magnitude of the effect bears little relationship to the change in readability formula scores.

Strategies exist for improving, early in the TM development cycle, the comprehensibility of textual materials. Specifications have a necessary but not a sufficient impact on comprehensibility. What is needed most, however, is a modification of quality assurance review cycles to include reviewers whose sole function is ensuring that TM materials are appropriate for the users.

#### Recommendations

1. The use of readability formulas to assess the difficulty of existing texts and to determine literacy gaps should be discontinued.
2. If predictive readability formulas are required, they should be developed in contexts similar to the ones in which they are to be applied. The predictor variables should be extended beyond words and sentences and even beyond the text itself, as may be necessary to reflect all contextual variables determining comprehension.
3. The use of readability formulas to regulate or guide the production of text should be discontinued.
4. The Navy should evaluate means of changing the management of text production to ensure more usable manuals.



## CONTENTS

	Page
INTRODUCTION . . . . .	1
Problem . . . . .	1
Purpose . . . . .	1
Background . . . . .	1
USING READABILITY FORMULAS TO PREDICT LITERACY GAPS . . . . .	3
Defining "Prediction" . . . . .	3
The Issue of Generality . . . . .	4
Formula Development . . . . .	4
New Formulas For New Applications. . . . .	4
A Military Example . . . . .	5
Rank-ordering vs. RGL-prediction Functions fo Readability Formulas . . . . .	5
Limits on Generalization . . . . .	6
Conditions of Reading . . . . .	6
Comprehension Measures. . . . .	7
Definitions of Comprehension used in the Development of Readability	
Formulas . . . . .	7
Comprehension Measures and Criteria . . . . .	8
A Military Example . . . . .	8
Valid Applications. . . . .	9
USING READABILITY FORMULAS TO GUIDE PRODUCTION OF TEXT . . . . .	9
A Readable Writing Strategy . . . . .	9
Writing to the Formula . . . . .	10
Writing Guidelines . . . . .	11
Readable Writing Research. . . . .	12
"Simplifying" Words or Sentences . . . . .	12
A General Writing Approach . . . . .	13
Marginal Benefits of Simplification . . . . .	13
Marginal Effects of Rewriting Text to Reduce Literacy Gaps . . . . .	15
CONCLUSIONS . . . . .	16
Predicting Reading Requirements . . . . .	16
Producing Comprehensible Text . . . . .	17
RECOMMENDATIONS . . . . .	18
REFERENCES . . . . .	19
DISTRIBUTION LIST . . . . .	23

## INTRODUCTION

### Problem

The armed forces have turned increasingly to the use of readability formulas to predict the ease with which users will be able to understand the text in technical manuals (TMs). Readability formulas are inexpensive, objective, and easy to use. They are, however, only proxies for the direct measurement of comprehension, and they have the attendant weaknesses of proxies. The validity of reading grade level (RGL) scores obtained by these formulas is questionable, as is the use of these scores in determining the usability of military TMs.

### Purpose

The purpose of this work was to review and evaluate the use of readability formulas by the military.

### Background

The volume of technical documentation has grown as the varieties and, perhaps more importantly, the sophistication of military equipment have increased. Muller (1976), plotting the growth of documentation for naval aircraft, points out that only 1800 pages were required to document the operation and maintenance of the Cougar Aircraft introduced in 1950. By 1975, 260,000 pages of TMs were required to document the F-14 fighter--a growth of 14,000 percent. The Navy now has an estimated 25 million pages of TM documentation and adds or revises 400,000 pages yearly (Sulit & Fuller, 1976). The Air Force spends an estimated \$70 million a year to add new manuals or revise existing ones (General Accounting Office, 1979). Across the services, there are 131,000 aviation maintenance manuals containing about 13 million pages (General Accounting Office, 1979).

This growth in TMs has not been accompanied by a comparable growth in the number of military personnel. Aiken (1980) compared the growth from 1946 to 1979 in the equipments and manning of Navy destroyers. While manning decreased by 9 percent over this time span, the number of equipments increased by 112 percent and the number of reparable parts increased by 600 percent.

Since TMs are the primary source of documentation for military equipment, it is critical to both military readiness and to individual safety that they be easy to use. However, the mushrooming of documentation has resulted in numerous problems as illustrated in a recent Government Accounting Office (GAO, 1979) report. Perhaps the most dramatic example in the GAO report involves the isolation and repair of one particular C-141 radar malfunction. It requires that the technician refer to 165 pages located in 41 different places in 8 separate documents. These disastrous effects of technological growth on the usability of documentation have been compounded by the low reading skills of many military personnel (Duffy & Nugent, 1978).

In designing a usable TM, four factors must be considered: access, accuracy, completeness, and comprehensibility. By access is meant the ease with which the technician can find the particular page or section required for the job at hand; the example of the C-141 radar described above is an example of poor access. Ease of access is largely a function of organization and indexing. The choice of an access system will vary as a function of user skill and literacy (Booher, 1978). However, once an access system has been selected for a TM, it can be concretely specified and the success of the

system can be evaluated without reference to the user. Once the technician accesses the relevant section, it is essential that all the information required for the job be present and accurate. However, even if the information is accurately presented, it will be of little use unless it is presented in clear and understandable manner; that is, unless it is comprehensible.

The Navy has initiated a major research and development effort to improve the usability of TMs (Sulit & Fuller, 1976). In this work, as in previous research, the criteria and procedures for ensuring comprehensibility have proven to be most elusive. In large measure, this is due to the fact that comprehensibility is inextricably tied to the interaction of the particular user with the manual. Access, accuracy, and completeness are very concrete characteristics of a manual. The user need not even be referenced in discussing these variables, except perhaps to note the gross skill level of the reader (apprentice vs. journeyman) when judging the degree of completeness required.

In contrast, a concrete, well-documented system for achieving a particular comprehension level is not available. A standard for comprehension cannot be specified without reference to the users and their interactions with the TM. A given TM is comprehensible if the users can understand and apply the information in it. The comprehensibility of any particular manual will vary as a function of the reading skill, the graphic interpretive skill, and the technical knowledge of the user. It will also vary with a variety of transient situational variables.

The judgment of whether or not a manual is comprehensible must therefore depend on a user test; that is, on the ability of a sample of users to perform a variety of tasks using the manual. Such a criterion has been specified in the acceptance standards for large numbers of TMs produced by firms under military contract. However, the actual evaluation is seldom carried out because of the expense and logistics involved. Valid tests would have to be developed first, and a sample of personnel would have to be gathered from each technical area relevant to the tasks described in the manual. These personnel would then have to be tested on a sample of job tasks from the text. Such a procedure is clearly unmanageable, given the high rate of TM production. As an alternative, the military has turned increasingly to the use of readability formulas.

Readability formulas are regression equations designed to predict comprehension (Klare, 1963). The predictor variables, typically, are word and sentence characteristics. In most instances, the formulas are based on the assessment of comprehension scores obtained from large samples of people reading selected passages. Some formulas yield scores that are the predicted years of education required to comprehend the manual (e.g., Flesch, 1949). Most formulas, however, yield a reading grade level (RGL) score that is the predicted reading skill required to comprehend the manual.

The formulas are seen as low cost and objective proxy measures for the actual assessment of comprehension; simply count the instances in which predictor variables occur, plug the numbers into the formula, and obtain a predicted comprehensibility score for the TM--or so it is thought. In industry as well as the military, readability formulas have been used to (1) determine whether "literacy" problems exist, (2) identify the areas where the problems are most severe, and (3) serve as a standards and specifications for the production of manuals. The first two uses correspond to the prediction function described by Klare (1976, 1979); the third use corresponds to Klare's production function. Klare (1979) has argued that readability formulas, while not ideal, are considerably better than other available instruments and have proven to be excellent tools for prediction in many situations. He further argues that several formulas have sufficient validity to be effective tools in guiding production.

In the following sections of this report, the author will argue that readability formulas have a very limited capability for predicting the comprehension requirements of technical documents. The regulations requiring the use of readability formulas assume that they are highly refined psychometric instruments that can be used to make point predictions of the level of comprehension to be expected; that is, the regulations call for material to be written at specific levels of difficulty based on the reading skill of the audience. Thus, the formula must be used not just to say that one manual is more difficult to read than another, but to assign to each manual a specific point on the reading grade level scale. The argument to be made in this report is that readability formulas are not designed to make these point predictions in any standard reading context. Rather, they can only be used to predict the relative difficulty of different texts.

It will also be argued that readability formulas are not effective production tools for ensuring that text is comprehensible. Further, a review of text production research indicates that the use of formulas as guidelines for rewriting does not result in practical improvements in comprehension.

## **USING READABILITY FORMULAS TO PREDICT LITERACY GAPS**

### **Defining "Prediction"**

"Prediction of readable writing" refers to the ability of a formula to assign accurate comprehension-difficulty scores to a large number of different passages (Klare, 1976). But what is an "accurate" score? In a weak sense, a formula is accurate if it can rank-order TMs in terms of their difficulty; that is, if it can predict that manual "A" will be more difficult to comprehend than manual "B." Note that there is no reference to the skill of the reader in this use except for the implicit assumption that manuals A and B are to be read by the same individuals.

In the strong sense, and in the vast majority of uses, "accuracy" refers to the extent to which the formula identifies the exact level of skill that would be required by a user of the document. For example, Dale and Chall (1948) state that the RGL score from their formula indicates the reading grade at which a book or article can be read with understanding. It is in this strong predictive sense that formulas have been most often used in industry and the military. Biersner (1975), for example, using a readability formula, found Navy TMs were written at an average of 14th grade level and assumed that this was too difficult for Navy users who, on the average, read at the 10th grade level. Duffy and Nugent (1978), Mackovak (1974), and Caylor, Sticht, Fox, and Ford (1973) all compared the readability formula scores of TMs to the reading skills of the readers to determine if there were "literacy gaps" that could affect job performance or learning; that is, whether the formula RGL score for the text was higher than the reading-test RGL of the users. They found significant numbers of personnel with a literacy gap as defined by such a score comparison and concluded that such gaps were likely to reduce the ability of these personnel to use their manuals effectively. However, as will be shown, these types of readability comparisons are of questionable validity and conclusions drawn from them are likely to be very misleading.

Kern (1979) has argued effectively that existing readability formulas are unsuitable for achieving the objective of matching the comprehensibility (or readability) of the text to the reader. His argument is based on an analysis of errors in the prediction of cloze comprehension when the FORCAST and Kincaid-Flesch formulas were applied to new materials. Absolute errors ranging up to nine grade levels were obtained when the readability formula scores for passages from military texts (other than the ones on which

the formula was based) were compared to the tested cloze comprehension. The error in prediction far exceeded the standard error value of about 1.6 grade levels. Thus, the two formulas did not produce exact predictions of cloze comprehension even under favorable conditions. While Kern's (1979) findings seriously question the validity of the formulas, the error may in fact be due to the small number of passages used in the development of most formulas; that is, the regression analyses on which the formulas are based involved no more than 20 sets of scores (20 passages). The error reported by Kern could be greatly reduced by larger and more divergent sampling techniques in the formula development procedure.

### The Issue of Generality

The argument to be made in this report extends beyond Kern's error analysis in that we contend that the formulas as presently conceived cannot be used, in principle, to predict the reading comprehension skills required to use a text on the job or in training. Exact prediction is impossible simply because the task being predicted, i.e., the task used in the development of the formula, is grossly different from the practical tasks for which TMs are used. On the other hand, the readability formula index has been related to a wide variety of indices of comprehension and a wide variety of comprehension tasks. Indeed, one could only wish that other experimentally obtained relationships would generalize as widely as has the readability work. However, it is not the determination of rank order relationships that is of concern, nor is this the use to which the formula is typically put. The concern is with the use of the formula to make exact predictions of reading requirements without adequately considering the effects of deviations from the conditions of development upon the accuracy of such exact predictions.

### Formula Development

Before examining the applications of formulas, the basic procedures and conditions in developing a formula will be reviewed. The development of most recent formulas has followed the same basic procedure (see, for example, Caylor et al., 1973; Kincaid, Fishburne, Rogers, & Chissom, 1975). First, comprehension of a set of passages is tested using a sample of readers with known reading skill. Each passage is then given an RGL score based on the RGL of the readers who comprehend the passages. Next, the instances of a variety of word and sentence features of each passage are counted (e.g., the number of letters and syllables per word; the number of words, prepositions, nouns, and phrases per sentence, etc.). An assessment is then made of the extent to which variations across passages in the numbers of the word and sentence features are related to variations in comprehension. Finally, the most strongly related features are entered into a regression analysis to develop the best linear prediction of the comprehension score for the passages. Most researchers find that a word factor (e.g., number of syllables per word) and a sentence factor (e.g., number of words per sentence) together yield the best prediction of the comprehension score (Entin & Klare, 1978). Thus, most readability formulas are of the following form,

$$\text{RGL} = a + b (\text{word measure}) + c (\text{sentence measure}),$$

where the expected RGL requirement is a function of an intercept (the constant "a") plus the sum of the weighted word and sentence factors.

### New Formulas For New Applications

Klare (1979) has counted over 100 different readability formulas. Given that they all follow the same general development strategy, why is there a need for so many formulas?

In some cases, the alternative formulas were developed to offer a choice between simplicity of application (a few, easily counted predictors) and accuracy of prediction (all predictors necessary to achieve the highest multiple correlation). Most formula development efforts, however, have stemmed from a concern over generalization; that is, whether a particular formula would be accurate in a given application, if the conditions of reading being predicted were generally different from the conditions under which the formula was developed.

Invariably, the focus is on either the similarity of the readers or the similarity of the passages in development and application. Can, for example, a formula developed using general literature passages be used to predict comprehension levels for school science texts?

### A Military Example

Just such an issue of applicability led each of the military services to develop its own readability formula (Caylor et al., 1973; Kincaid et al., 1975; Smith & Kincaid, 1960). The military formulas were developed because it was felt that the requirements of a military technician reading a TM would not be predicted accurately by formulas based on children reading children's textbooks, such as the Dale-Chall (Dale & Chall, 1948) and the Flesch Reading Ease (Flesch, 1948) indices. The appropriateness of word and sentence factors as predictors was not questioned, nor was the comprehension criterion (though the latter was changed by necessity). Rather, the goal was simply to get new values for the intercept and the weights in the basic formula. Kincaid et al. (1975) report that they sought simply to recalculate three existing formulas using Navy personnel and materials.

Formulas based upon texts from elementary and secondary schools were felt to predict too high a reading skill requirement for two reasons. First, long technical words in TMs are familiar to technical readers, but are nonetheless scored as difficult in the school-based formulas. Thus, renorming using Navy materials should result in a lowering of the weight given to the word-length factor. Second, it is thought that an adult with an RGL of 9.0 will actually comprehend more of a TM than would a child with the same RGL (see Curran, 1980). The effect of norming using Navy personnel would thus be to decrease the size of the intercept (a) in the basic readability formula given on page four.

### Rank-ordering vs. RGL-prediction Functions of Readability Formulas

In summary, it should be clear from the preceding discussion that new formulas were developed with the intent of increasing accuracy of prediction in the strong sense; that is, they were developed to ensure that the score resulting from the application of the formula could be referenced to the skill required of the user.

New formulas would not have been required for simply rank ordering the difficulty of the manuals; the military formulas and the school-based formulas are highly correlated. Caylor et al. (1973), for example, obtained correlation coefficients of .94 between the Army's FORCAST formula and the Dale-Chall and Flesch Reading Ease formulas, respectively. Thus, the rank ordering of TMs by difficulty could be done as well with a school-based formula as with any of the formulas used by the military. In fact, Klare and Smart (1973) found a formula based on the performance of children was highly effective in predicting the relative difficulty of military correspondence manuals. The formulas used by the military, however, were developed to make exact grade level predictions of reading requirements, and that is how they have been used.

### Limits on Generalization

Most of the literacy gap research discussed thus far (Biersner, 1975; Caylor et al., 1973; Duffy & Nugent, 1978) was based upon formulas developed for the military. This research attempted to predict the reading skill required to use technical materials regardless of whether they were for use on the job, in correspondence training, in classroom instruction, or in self study. Because these formulas were based on studies of military men and materials, it has been assumed that they should be able to predict, within the standard error of estimate, the exact RGL a user must have to comprehend any given manual regardless of the comprehension task or the situation in which it is carried out. Thus, a given formula may be expected to predict a variety of reading-to-do and reading-to-learn tasks (Sticht, Fox, Hawke, & Zapf, 1977). These military formulas are used not just to say that manual A will be easier to read than manual B but also to say that manual A will require, for example, a 10th grade or better level of reading skill to be used effectively.

In generalizing readability formulas, researchers have forgotten half of the development particulars. The development of a formula involves not only a particular set of people reading a particular set of passages but also the assessment of a particular type of comprehension under particular reading conditions. In the same way that there have been reservations in generalizing to different readers and to different passages, so there must be concern about generalizing to different comprehension tasks and reading conditions. These situational variables are no less important in specifying the limits for generalizing a formula than are considerations of the particular texts and the particular readers used in the development.

Klare (1963), in discussing the limitations of readability formulas, states that they do not measure the effects of the user's purpose in reading or the effects of format, typography, or content. It is quite true that these variables were not manipulated in the development of the formula and therefore the formulas do not reflect changes in the variables. However, the passages used must have had some format and the reader some purpose. The time for reading and the nature of the questions had to be specified; that is, the developer, while not manipulating these variables, had to fix them at some value. In using the formula, predictions of comprehension will be in error to the extent that the developers' assumptions are violated.

### Conditions of Reading

Consider the effects of just a few of these variables. In developing a formula, the readers are subjects in an experiment and thus not very well motivated. Suppose, however, that these subjects were told that their promotions depended on their comprehension scores. Scores would zoom up and, given a fixed comprehension criterion, the resulting readability formula would predict all manuals to be much easier. If the application of the formula was to be for manuals used in studying for promotion, then just such motivation instructions should be given in development; that is, if accurate prediction is the objective. Similarly, allowing the subjects two or three readings of the text, as occurs in typical studying, will result in higher scores and, if the criterion for comprehension is fixed, predictions of higher readability. Using smaller typefaces, such as those found in many manuals, will tend to reduce comprehension if reading times are restricted.

Reading time affects comprehension scores. Klare (1976, 1979) has stated that readability formulas are not predictive when reading time is unlimited. However, his focus is on the relative difficulty of materials (i.e., the weak prediction of accuracy). If

we want to predict the level of reading skill required to achieve a specified level of comprehension (e.g., 75% correct on a factual multiple-choice test), then the accuracy of our predictions will depend on (1) the time allowed for reading when the formula was developed and (2) the time allowed in the situation for which the prediction is being made. Obviously, a given formula cannot predict the particular reading skill required to comprehend a book irrespective of whether the reading time allowance is 100, 200, or 300 words per minute.

These situational variables can have a major effect on predictions of RGL requirements. Yet, it is highly unlikely that more than a small minority of the situational variables encountered in the development of a formula will match the situational variables found in its application. Thus, it would be inappropriate to make exact predictions using the formula.

### Comprehension Measures

Of even greater significance than the situational variables are the definitions of comprehension used in developing formulas. The grade level score from a formula is not the grade level required for some amorphous, universal comprehension task. It is the grade level required to accomplish a very specific comprehension task to a very specific criterion level. If the skill requirement is to be predicted for a different reading task or a different level of performance, then the effects of that change on performance must be known and included as a variable in the readability formula. Failure to do so will almost certainly contribute to spurious predictions.

### Definitions of Comprehension Used in the Development of Readability Formulas

Consider the measures of comprehension used in the military formulas. Kincaid et al. (1975), in developing the Kincaid-Flesch formula, assigned comprehension scores to passages based on a combination of performance on a cloze test (Taylor, 1953) and performance on the Gates-MacGinitie reading test (Gates & MacGinitie, 1965). Specifically, an individual was said to comprehend a passage if he or she scored 35 percent or more on a cloze test of that passage. The reading grade level required for comprehension of the passage was then determined by first categorizing the readers into RGL categories (i.e., readers with RGLs of 8.5 to 9.4, 9.5 to 10.4, etc., based on the Gates-MacGinitie test). Each group was then examined to determine if 50 percent or more of the readers in that group comprehended that passage (i.e., scored 35 percent or better on the cloze test). The passage was assigned the RGL of the lowest Gates-MacGinitie RGL group meeting the criterion. Thus, if a TM has an RGL score of 10.0 on the Kincaid-Flesch formula, it means that at least 50 percent of the readers with an RGL of 10.0 on the Gates-MacGinitie test may be expected to score at least 35 percent on a cloze test of the TM.

What does this comprehension test and comprehension criterion have to do with the skill required in reading to do a job or reading to pass a test? Compare this reading criterion with the definition of comprehension in a self-study course such as the Navy's Basic Electricity and Electronics Course. In that course, comprehension is defined as a score of 100 percent on a closed-book multiple-choice test taken after the student has spent no more than the allotted number of hours or days studying the chapter and receiving information clarifying the text when requested. In correspondence courses, there is a different criterion and it is generally an open book test. In advanced training, there are lecture supplements and there is generally an open book test.

How is a tenth grade readability score based on 35 percent cloze comprehension to be interpreted in judging the appropriateness of a TM for the personnel who have these



actual comprehension requirements? One might argue that, since our concern is the comprehensibility of the text, the effects of lecture supplements will have to be ignored in developing the criteria. But, if the formula is to be used to predict comprehension requirements under real world conditions, it is of little use to develop a formula to predict the skill requirement needed to understand the material when reading it in isolation.

### Comprehension Measures and Criteria

The comprehension measures and criteria used in most readability formulas are quite arbitrary. Why not score synonyms correct on the cloze tests instead of requiring the exact word that was deleted? The correlation would not change but the predicted comprehensibility of the passages would. Reducing the percentage of people required to demonstrate comprehension in an RGL category from 75 to 50 percent sounds like a minor and rather arbitrary decision. Who could say which was the "proper" criterion? Yet such a reduction could result in a 3 or 4 grade level change in predictions made using the resulting formula. Since the decisions made in establishing comprehension criteria are arbitrary, the resulting predictions of comprehension requirements must also be arbitrary (again, in the absolute sense).

### A Military Example

The arbitrariness of readability formula scores can perhaps be illustrated most clearly by an examination of the assumptions and errors made in establishing the comprehension criterion for the Army's FORCAST (Caylor et al., 1973) and the Navy's Kincaid-Flesch (Kincaid et al., 1975) formulas. As noted previously, a 35 percent cloze criterion was used in both efforts with the expectations that 35 percent cloze was equivalent to a 70 percent multiple-choice score.

This equivalency was assumed on the basis of the authors' interpretations of two reports on the relationship between multiple-choice and cloze testing (Bormuth, 1967; Rankin & Culhane, 1969). There are two problems with this criterion. First, a 70 percent multiple-choice comprehension score on a passage is taken by reading teachers to indicate that the reader is at the "instructional level" in attempting to comprehend the passage; that is, the reader cannot adequately comprehend the passage without assistance (Entin & Klare, 1978). This would obviously result in an inadequate match of reader to TM if the TM was to be used on the job or in independent study.

For military use, readers should be able to read and comprehend the TM independently. Reading teachers consider a multiple-choice comprehension score of 90 percent to reflect this criterion (Entin & Klare, 1978). A multiple-choice score of 90 percent has been found by Bormuth (1967) to equate to a 50 percent cloze score. Thus, the military readability formulas should have been based on a 50 percent cloze criterion instead of a 35 percent cloze if the authors wanted the formulas to predict the reading skill to work independently with the manual.

If the "instructional level" of comprehension was the goal of prediction in the military formulas, even this goal was not achieved. This is because the second problem with the criterion is that a 35 percent cloze score does not, in fact, equate to even the instructional level represented by a multiple-choice score of 70 percent. As Klare (1979) pointed out, the findings of Bormuth (1967) and Rankin and Culhane (1969) were misinterpreted in developing both the FORCAST and Kincaid-Flesch formulas. Bormuth (1967) and Rankin and Culhane (1969) found a 40 percent cloze score equated to a 70 percent multiple-choice score. Klare (1979) estimated that a 35 percent cloze equated to only a 50 percent multiple-choice score. Thus, the comprehension criterion set by Kincaid

et al. (1975) and Caylor et al. (1973) is not only below the instructional reading level, it is even below the "frustration" level (i.e., the level at which a reader will quit reading).

Even if readability formula scores were not subject to the problems discussed earlier, the matching of users and manuals by using these military formulas throughout the services would result in frustration in reading and rejection of all the manuals by the readers. Thus, if the scores had any absolute meaning, the Kincaid-Flesch and FORCAST formulas would have to be withdrawn and renormed.

### Valid Applications

Because of the arbitrariness of readability scores, they have very limited practical use. Two uses come to mind. First, the formulas may be used to choose between alternative TMs for specific groups of readers. The aim may be to select the easiest manual for the group, or, if the text materials for a rating are to be revised, a formula may be used to identify the most difficult so that they may be revised first.

Second, the readability formula score could be used as a variable in relating comprehensibility to another variable. For example, Klare and Smart (1973) found the readability score of military correspondence course manuals correlated .75 with course attrition. Other work has included readability as an independent variable in factorial experiments on text comprehension (Klare, 1979). In all valid applications, however, the concern is relative--not absolute--difficulty.

## **USING READABILITY FORMULAS TO GUIDE PRODUCTION OF TEXT**

In prediction, readability formulas have been used to assess the comprehensibility of already-written materials; that is, to identify text in use that is likely to be difficult to comprehend. Obviously, readability formulas would be of much greater value if they could be used at the time the text is written. The military and other large organizations (Redish, 1979) are beginning to use readability formulas in just this way (see Curran, 1977 & 1980; Kern, 1979; Department of the Air Force, 1977; Department of the Army, 1978; Pressman, 1979).

The basic application of readability formulas in production is as feedback to the writer. In some cases, where computer editing systems are used, the feedback may be provided after each paragraph is written. This feedback would include not only the formula readability score for the paragraph but also identification of the particular words and sentences that were judged "difficult" (Curran, 1977; Kincaid et al., 1975). The feedback may simply serve as guidance to the writer, in which case the writer can examine the material judged difficult and make a personal determination as to whether or not changes are required (the readability assessment was valid). Of course, if the writer accepts the validity of the formula score, it is incumbent upon him or her to rewrite the materials until an acceptable score is obtained.

Most typically, the readability score is used not only as feedback but as a criterion that must be met. Thus, the writer must rewrite the text whenever the required readability score is exceeded.

### A Readable Writing Strategy

If the readability formula score is a required criterion, or if the writer accepts the predictive accuracy of the formula score, "difficult" passages will have to be revised to

improve their readability. Klare (1979, pp. 82-83) provides a step-by-step procedure for the proper use of readability formulas in carrying out such revisions:

1. Apply a formula to see if a piece of writing is likely to be readable to intended readers.
2. If the readability index suggests it is and if other requirements for good writing have been met, stop there. Keep in mind, in other words, that, while a poor index value predicts poor writing, a good index value by itself need not mean good writing.
3. If the readability index suggests the piece of writing is not likely to be readable to intended readers, put the formula aside so as not to be tempted to "write to formula."
4. Rewrite the material, trying to discover and change those parts likely to cause trouble. Use the formula information only as a guide as to where to begin.
5. Apply the formula again, to see if the piece of writing is now more likely to be readable to intended readers.
6. If it is, and other requirements for good writing are met, stop there.
7. If it is not, repeat steps three, four, and five until an appropriate readability index is achieved.

Klare's procedure raises two questions: (1) can the writer "put the formula aside" and (2) what are the clear writing techniques that will both improve comprehensibility and reduce the formula score?

#### Writing to the Formula

Step three in Klare's procedure contains the critical requirement that the writer put the formula aside while rewriting. This means that the writer must not "write to the formula" by changing only those variables indexed by the formula without considering whether or not the change will make the material easier to understand.

As Klare (1979) points out, all experts in the field of readability agree that writing to formula is ineffective.

Can we, however, really expect the writer to put the formula aside if the objective is to improve comprehension and reduce the readability score? One might expect that a writer will be better able to set the formula aside to the extent that other comprehension criteria are available. For example, in newspaper and magazine writing, the real comprehension criterion is readership--if your writing does not attract readers, you will be fired even if your articles achieve low formula scores. A writer in this situation would be foolhardy to write to the formula.

But what of the person writing to a military specification that includes as its only criterion for text comprehensibility the achievement of a specific grade level score? Suppose that a writer prepares a draft TM that he or she considers to be complete, accurate, and comprehensible. If the criterion from the formula is not achieved, and if the formula score is the only contractual criterion for the acceptability of the text, then the writer will be forced to rewrite to the formula, regardless of the effect this will have on the usefulness of the TM. The tendency will increase to the extent that the writer's

job depends on meeting budgetary constraints and tight production schedules, both of which are common circumstances in TM production (Duffy, 1982).

Even if the money and time constraints are minimal, writers will tend to write to the formula to the extent that they find it difficult to otherwise achieve the readability criterion. Hooke, De Leo, and Slaughter (1979) reported that Air Force writers found it extremely difficult to achieve readability scores below an RGL of 10.0. In sum, it is likely that technical writers will, in a significant number of instances, write to the formula if it is possible to do so.

If readability formulas are to have any possible effectiveness as production criteria, then they must be designed so that they cannot be "written to." Typically, formulas are generated for ease of application, and variables that are highly correlated with the primary predictors are dropped from the formula. However, for use as a production criterion, the highly correlated variables could be left in the formula, thus increasing the number of predictors and making it difficult to write to the formula. Bormuth (1969) has developed formulas with up to 24 predictor variables. While it would be virtually impossible to write to such formulas, they could be programmed for easy computer application and they would reflect the effects of revisions that would be expected to improve ease of understanding.

### Writing Guidelines

If the writer is able to put the formula aside, what writing techniques should be used to revise the text? Step five in Klare's (1979) procedure calls for the reapplication of the formula after the revision is complete. Further, if the text is still scored as too difficult, the formula is to be put aside again while a second revision is made. The formula is then reapplied. If the formula is to be effective in this iterative process, it is essential that the meaning of a formula score is the same when the formula is applied to the last draft as it was when applied to the first draft. For this to occur, the revision effort must have an equivalent effect on both the readability score and comprehension; that is, the linear regression relating the readability variable and actual comprehension must remain constant. Writers must be able to interpret the formula output in the same way from application to application.

In essence, then, while one does not write to formula, it is essential that whatever changes are made will affect the variables indexed by the formula. Since even the most complex formulas are restricted to the measurement of sentence and word characteristics, production guidelines must obviously focus on the simplification of words and sentences. In his Manual for Readable Writing, Klare (1975) describes the process of making writing more readable as "changing words" and "changing sentences" to make them easier to understand. Graphics, format, and organization, while important to comprehension, are not a part of the readable writing process. Simplifying graphics, for example, will not reduce the formula score, and hence is irrelevant to the revision process, when a formula is used for feedback.

There are innumerable style manuals available, and each recommends a variety of techniques for improving comprehension. Included in the recommendations are techniques for readable writing (e.g., simplifying words and sentences). Klare (1963, 1975, & 1979) and Flesch (1949) specifically address readable writing and present guidelines for improving both readability and comprehension. Their recommendations include the manipulation of word dimensions (e.g., familiarity, concreteness, and association value) and grammatical class (e.g., increasing the proportion of function words and avoiding nominalizations). Sentence recommendations include using active sentences with few

dependent phrases (embedding). It can be easily demonstrated that following these guidelines will improve readability, at least as it is indexed by most formulas. Active sentences tend to be short sentences, whereas embedding lengthens sentences. Familiar words are usually short words. However, the guidelines must improve comprehension as well as readability, and here the evidence is not so clear.

### Readable Writing Research

There has been very little research on the effects of these variables on comprehension; hence, guideline recommendations have been inferred from verbal learning research (Klare, 1975). The verbal learning research, however, has been on the learning of word and sentence lists and on the verbatim recall or recognition of those lists. Surprisingly little is known about the generalization of list-learning research to text comprehension (Goetz, 1975). Where tests have been carried out, the generalizations have been difficult to specify (Reder, 1978).

Klare (1976) was able to identify only 36 studies since the mid-1940s that evaluated the effects of readable writing techniques on text comprehension. Readable writing variables were confounded with other variables in many of these studies, making valid evaluations impossible. In the extreme, the "readability" comparison was between passages from different books. Klare reported that there was "evidence of an attempt" to control content in only 11 studies. In some of the controlled studies, however, the text revision nonetheless involved considerably more than the application of readable writing guidelines. For example, Hiller's (1974) simplification of a 1200-word mathematics passage increased the length by 18 percent and included the addition of a concrete example. Feldman (1964) controlled content, but passage length increased by 40 percent. Obviously, more than sentence and word simplification was involved.

Very few of the studies reviewed by Klare (1976) or published subsequently have evaluated specific readable writing guidelines. The research that has been carried out, however, has failed, in the main, to find any effects of practical significance due to application of the guidelines.

### "Simplifying" Words or Sentences

Nolte (1937), in one of the earliest studies of the effects of applying readable writing techniques, simplified passages using the requirement that all words be on a fourth grade vocabulary list. Although an extensive test program was carried out, no effect on comprehension could be demonstrated.

Duffy and U'Ren (1982) also used vocabulary lists to simplify passages. Although 25 percent of the content words in eight passages were simplified, effects of practical significance were obtained only under very specific conditions and in only one of four experiments. Tuinman and Brady (1973) held the passages constant across conditions but "simplified" by teaching the unfamiliar vocabulary to the students in a series of sessions extending over a week. While the instruction improved vocabulary knowledge by 20 percent, there was no effect on comprehension.

A similar lack of significance has resulted when sentence variables have been manipulated. Duffy and U'Ren (1982) revised passages using a rule that every sentence must be a simple sentence with no adverbial or prepositional phrases. In a series of four experiments, sentence length was reduced from 20 words per sentence to 10, yet no comprehension effects were obtained. Coleman (1962) varied the average sentence length of a passage from 16 to 39 words by applying Flesch's (1949) readable writing guidelines

for sentences. While the results were statistically significant, the effects were meager and of little practical significance.

Coleman (1962) concluded that shortening sentences may not be an effective readable writing strategy. In a post hoc analysis, he examined three simplification strategies. Breaking a compound sentence joined by "and" into two sentences had no effect on comprehension. Raising clause fragments (e.g., participial, gerundial, and infinitive phrases) in a complex sentence to the status of full sentences resulted in only marginally significant improvement in comprehension. Only breaking sentences joined by coordinate conjunctions other than "and" resulted in an improvement in comprehension that was likely to be reliable.

### A General Writing Approach

The research reviewed thus far has focused on the evaluation of specific guidelines for simplifying vocabulary or sentences. The findings have failed to support the validity of any specific revision strategy as a means of improving comprehension. It has been argued, however, that the readable writing approach cannot be validated by the separate validation of individual guidelines (Klare, 1976; Nolte, 1937). Indeed, Klare (1976) has suggested that the manipulation cannot even be restricted to just vocabulary or to just sentence simplification. The argument is that the piecemeal application of individual guidelines will result in awkward, stilted writing, thereby counteracting the effects of simplification.

Thus, it is argued that the test of the validity of readable writing strategies must involve the application of a general readable writing approach that involves simplifying both sentences and words. A style manual or checklist might be used to provide writers with such an approach.

Developing and validating a general readable writing approach is fraught with difficulties. Such an approach must be made up of a series of individual guidelines. Yet, since individual guidelines cannot be validated, there is no empirical way of determining which guidelines were effective and hence which should remain in the general approach. Without the ability to validate individual guidelines in some way, it is quite likely that a general writing approach will include guidance that is ineffective (e.g., Coleman, 1962) or even detrimental (e.g., Pearson, 1974-1975) to comprehension. Thus, a significant part of a revision effort based on such a general approach could be counterproductive.

### Marginal Benefits of Simplification

Klare (1976) judged findings from evaluation of general readable writing approaches on the basis of the statistical significance of the effects and concluded that readability makes a difference, sometimes. The present author must add that readability makes a practical difference, seldom. Inconsistencies caused by ineffective guidelines may account for the fact that effects are weak at best, even when a general approach to readable writing is used.

Kincaid and Delionbach (1973), in one of the few statistically significant studies, rewrote passages from a military maintenance manual to the 8th, 12th, and 16th grade levels. The eighth grade manipulation resulted in an increase of only 7 percentage points on a multiple-choice comprehension test. There was no difference in performance between the 8th and 12th grade versions. Klare, Mabry, and Gustafson (1955) obtained a statistically significant improvement of 8 percent in multiple-choice performance when a 16+ grade level version of a military maintenance passage was simplified to the 7th-8th

grade level. A middle version was not significantly different from either of the extreme versions. One might argue that a reliable 7 to 8 percent gain in performance is of practical significance. However, the gain is quite small, one percentage point per grade, relative to the effort required to make an eighth-grade or greater reduction in the RGL. Additionally, writers seldom miss their target RGLs by eight grades. The more likely event is that the writer will overshoot by four or five grade levels at the most. With this smaller degree of simplification, neither Kincaid and Delionbach (1973) nor Klare et al. (1955) achieved reliable improvements in performance.

In studies by Klare et al. (1955), Kincaid and Delionbach (1973), and in most of the other readable writing studies, neither the motivation nor the reading skill of the readers was assessed or controlled. Klare has argued that the failure to control such variables may account for the many weak and nonsignificant effects. Klare (1976) has presented a model of the reading situation as it applies to readability in which it is argued that the failure to control these and other variables may account for the many weak and nonsignificant effects. Basically, if the reader is well motivated and has sufficient reading time, he or she will be able to work through a text regardless of style difficulty. Similarly, if the reader is already very familiar with the topic or if the readability scores for all versions of a passage are either all above or all below the reading skill of the users, the manipulation of the text cannot be expected to have more than a minimal effect on comprehension.

Fass and Schumacher (1978) have extended Klare's (1976) model to include the reader's processing activity as a critical intervening variable. They propose that difficult text requires more elaborate or deeper processing than simple text. If the reader can and does engage in appropriate processing activities, then the effects of simplification will be negated. Their argument is in the context of the learning of a lengthy passage but the argument may be expected to apply to any comprehension task requiring inference or long-term memory. Many of the variables in Klare's (1976) model (e.g., motivation, background knowledge, and reading time) can be interpreted in terms of their effect of processing activity.

Two recent studies carefully attended to the variables identified in Klare's model, yet failed to offer any support for the use of readability formulas as production guidelines (Duffy & U'Ren, 1982; Kniffen, Stevenson, Klare, Entin, Slaughter & Hooke, 1979). However, the Duffy and U'Ren study did yield some evidence that processing requirements are relevant variables in determining the effectiveness of simplifying text. In both studies, the manipulation of readability failed to facilitate comprehension and none of the variables discussed by Klare (1976)-- motivation, reading time, nor difficulty level--could account for the lack of effects.

Kniffen et. al. (1979) manipulated the literacy gap--the difference between the readability score for the materials and the reading skill score for the readers. Two different 5000-word samples of technical materials were rewritten to RGLs of 8, 10, 12, and 14 using Klare's (1975) Manual for Readable Writing. These materials were then administered to military personnel with 8th and 10th grade reading skills to create literacy gaps of 0, 2, and 4 RGLs. Reading time was manipulated to allow reading rates of approximately 85, 130, or 175 words per minute. A carefully constructed multiple-choice test was administered after reading. In the first analysis, performance on each passage was analyzed separately. Neither the literacy gap effect nor the interaction of literacy gap with reading time was significant in either analysis. In a subsequent combined analysis, the literacy gap did produce a statistically significant effect: Comprehension test performance improved by five percentage points, hardly an effect of practical significance. Even in the overall analysis, the literacy gap did not interact with

reading time, thus failing to support the hypothesis that restricted reading time will enhance readability effects. Duffy and U'Ren (1982) simplified the eight passages in the Nelson-Denny Reading Test (Nelson & Denny, 1960) using a restricted vocabulary list (generally words at or below the fourth grade level) and a syntactic complexity limitation. Thus, both sentences and vocabulary were simplified using fundamental readable writing strategies. Every attempt was made to maintain a smooth writing style. The result of the manipulation was a reduction of average Kincaid-Flesch (Kincaid et al., 1975) readability from the 11.5 grade level to the 5.5 grade level.

The tests were conducted in a low motivation context. The readers were Navy recruits in the midst of basic training. No special incentives were provided for good performance, the testing was unrelated to their basic training, and they knew that their performance scores would be confidential. Thus, the conditions of motivation were such as to maximize the effects of the readability manipulation. A reading skill pretest was given to the subjects so that the interaction of reader skill level with the change in readability could be evaluated. Since reading skill levels varied from less than 7th grade to college level, a wide range of literacy gaps was evaluated in the interaction. Across the experiments, the researchers manipulated the comprehension test (cloze vs. multiple choice), reading time, and the memory requirement. There was an attempt to address, either within or between experiments, each of the major variables called out by Klare (1976) as moderators of the comprehension effect of readable writing manipulations. The only readability effect of practical significance was achieved by simplifying vocabulary for low ability readers when memory was required. In all other conditions, across all the experiments, the researchers failed to find practical effects of any of the readable writing manipulations. There was no trend toward a readability effect with decreasing reading skill or reading time. Even in the memory experiment, the effects were not consistent with readability predictions. In fact, the vocabulary simplification that resulted in the improved comprehension was the one that produced the smallest change in readability score.

#### Marginal Effects of Rewriting Text to Reduce Literacy Gaps

Simplifying both vocabulary and syntax consistently failed to facilitate comprehension. Thus, the findings offered no support for the use of readability formulas as feedback devices for predicting the effects of simplification. The fact, however, that the simplification effect was only obtained when the task involved a significant memory component offers some support for Foss and Schumacher's (1978) proposal that the effectiveness of simplification will depend on the processing demands of the task.

In summary, the findings of Duffy and U'Ren (1982), along with those of Kniffen et al. (1979), present a strong case against the readable writing approach to revision and hence against the use of readability formulas as feedback devices for writers. Duffy and U'Ren (1982) used fundamental readable writing techniques to rewrite materials from a widely used reading test. Kniffen et al. (1979) used a readable writing style manual. In both cases, conditions were optimal for the readability improvements to facilitate comprehension. Yet, in both cases, the manipulations, with one exception, resulted in no effect or, at best, a marginal effect on comprehension. If the revision approach does not produce large comprehension effects under these ideal testing conditions, then there must be little expectation for the approach to be effective in practical applications. In fact, the findings of Duffy and U'Ren (1982) suggest that some readable writing techniques will not be effective in improving comprehension under any circumstances. The effectiveness of other simplification strategies will depend on the reading requirements and reading conditions.



## CONCLUSIONS

The military has a well defined need to identify both the kinds and the levels of reading skills required to use its TMs effectively. While readability formulas have frequently been employed to specify these requirements, it must be concluded that the applications have not been valid. The lack of validity does not simply mean the lack of a scientific nicety. It means, in fact, that very large errors in prediction are being made, and, as a result, erroneous conclusions are being drawn about the difficulty personnel are experiencing with TMs. Depending on the comprehension task, these conclusions may underestimate the degree of difficulty as readily as overestimate it.

### Predicting Reading Requirements

Accurate prediction of reading requirements is most certainly possible. However, if the readability approach is to be used, then new readability formulas must be developed. Just as the military has sponsored the development of new formulas based on military personnel and TMs, it will have to sponsor the development of new formulas based on military comprehension tasks. Thus, there would be formulas to predict the comprehension requirements of correspondence courses, platform instruction, self-study materials, conceptual job tasks, and procedural job tasks. Formulas could be refined to the point (e.g., self-study with and without audiovisual support) where predicting reading requirements would give way to empirical assessment of each case. However, practical considerations would rule that out. The minimum requirement would be to at least represent the generic comprehension task being predicted. Once that was done, the validity of the readability formula could be checked by comparing predicted comprehension by different readers with the actual comprehension scores they obtained in school, on the job, etc.

Consider the development of such a formula for correspondence texts. The unit of analysis would not be a 200-word passage but, rather, the unit that is tested—probably the chapter. Comprehension of the chapter would be indicated by the score obtained on the regularly administered test. If items on this test were considered inadequate in number or quality, additional instructor-approved items would be generated and administered as a second test. Next, the ability of the test takers would be indexed using the Armed Services Vocational Aptitude Battery test scores. Either a reading test (word knowledge), a job relevant test, or a test composite (in standard score form) could be used.

In a typical readability formula, reader ability is taken into account in determining the criterion (e.g., can 50% of the readers at a particular RGL score 35% on the comprehension test?). In the proposed procedure, reader ability would be a predictor variable (i.e., comprehension for the particular manual-reader combination would be predicted). Next, physical attributes of the chapter would be indexed. Since a whole chapter is used, it is possible to go beyond word and sentence factors and include formatting and graphic factors. Finally, a predictor function would be generated of the form:

$$\text{Proportion correct} = a + b \left( \begin{matrix} \text{reader} \\ \text{skill} \end{matrix} \right) + c \left( \begin{matrix} \text{word} \\ \text{factor} \end{matrix} \right) + d \left( \begin{matrix} \text{sentence} \\ \text{factor} \end{matrix} \right) + e \left( \begin{matrix} \text{format} \\ \text{factor} \end{matrix} \right) + f \left( \begin{matrix} \text{graphic} \\ \text{factor} \end{matrix} \right).$$

An applied formula such as this may still not be effective because of excessive error variance. If this turns out to be the case and the error variance can not be accounted for, then the readability prediction effort should be abandoned. After all, the readability formula is an empirical tool, not a theoretical construct. If real world variations in comprehension cannot be predicted, it serves no purpose to turn to accurate predictions of

artificial situations. In fact, practical readability formulas should be able to predict practical comprehension as well as the existing military formulas predict cloze comprehension, but that may still be unacceptable (Kern, 1979).

An alternative to predicting the reading skill requirement of a manual is the prediction of job reading requirements. Here the manual is but one component of the job and, even then, only those parts of the manual actually used are considered. Sticht et al. (1977) carried out an exploratory investigation of this approach. Samples of job reading materials and tasks were obtained via interviews. Next, the reading tasks were scaled for the reading skill required for successful completion. These materials were then presented to Navy personnel in a survey to rate the frequency with which each reading task was carried out in the course of a week. A weighted average of the reading skill demands provided a statement of the reading requirements. Both text and graphic requirements can be specified using this procedure.

### Producing Comprehensible Text

The military produces an enormous number of pages of text annually. Procedures exist to ensure, within reasonable cost boundaries, that this text will be comprehensible to the users. The present author found that readability formula scores will not be effective as production criteria. Further, focusing on readable writing guidelines in revising text yields virtually no practical benefit to comprehensibility. However, revisions based on readable writing guidelines can be effective at the extreme levels of difficulty. If the reader has no knowledge of the meaning of a significant proportion of the vocabulary, the sentences are extremely complex, and the reading task is more than a "look up," then a readability formula score can be an effective criterion requirement for producing improvements in comprehension—that is, if the writer does not write to the formula.

In addition to being ineffective in most situations, the use of readability formulas seems to have limited the consideration of other comprehension factors. More than just sentence and word factors determine comprehensibility, especially in TMs. Graphics play an integral role in TMs, yet little attention is given to their design or to their coordination with text. Procedural listing vs. paragraph presentation of information, highlighting techniques, and the organization of information within paragraphs may all be expected to affect the comprehensibility of text.

How are all of the comprehensibility factors to be taken into account in the production of manuals? There are three alternatives: guidelines, regulations, and changing the production system.

1. Guidelines. Guidelines would not appear to be an effective approach. Most guidelines are quite easy to understand (e.g., place text and relevant graphic on the same or facing pages), yet they are violated constantly. There are innumerable books and training courses providing guidance for technical writers; yet, comprehensibility continues to be a problem. Thus, guidance alone has proven ineffective.

2. Regulation. The specification of readability formula scores as criteria for acceptance of technical materials is an attempt at regulation. Although the standard readability formula has not been effective, a more complex formula, one that included graphics, highlighting, and other comprehensibility factors, might be. The Standard for Comprehensible Writing (Department of Defense, 1978) attempts to translate all relevant research on comprehension into concrete writing and design statements. For example, the number of graphics per page, the use of procedural statements, and the use of specific highlighting techniques are all described in such detail that the standard could be used as

a specification for producing TMs. While the implementation of a specification of this complexity might increase comprehensibility, it probably would not, by itself, be cost effective. It would be necessary to train writers and designers in the use of the specification, and all details of each draft TM would have to be reviewed in relation to the explicit specifications. However, through a gradual evolution, including the development of training courses and the programming of the specifications into computer editing systems, a cost effective procedure for controlling the comprehensibility of manuals could be developed.

3. Changing The Production System. A similar but more flexible system for controlling comprehensibility is embodied in McDonald-Ross and Waller's (1976) concept of a "transformer" in the production process. The transformer is an individual or group whose sole responsibility is to ensure that the text (or TM) is maximally usable for the intended audience. The transformer has competence in educational technology, editing, graphic design, and the subject matter area. The transformer, then, has the responsibility for ensuring that the principles of good writing, such as those embodied in the Standard for Comprehensible Writing (Department of Defense, 1978) are applied appropriately.

An example of a transformer system can be found in the Navy's hardware procurement. The procurement of technical hardware has repeatedly encountered the same problem that recurs frequently in the procurement of technical documentation--the design process does not adequately attend to the manning requirements (i.e., the needs of the user). In an attempt to address this problem, the Navy has established an office, whose acronym is HARDMAN (Chief of Naval Operations, 1977), which has as its sole function the reviewing of each phase of the procurement effort to ensure that the "people considerations" are fully attended to. It is only through the institution of a complex specification or through the institution of a transformer office analogous to the HARDMAN office that all aspects of text relevant to comprehension can be controlled.

## RECOMMENDATIONS

1. The use of readability formulas to assess the difficulty of existing texts and to determine literacy gaps should be discontinued.
2. If predictive readability formulas are required, they should be developed in the same kind of context for which they are to be applied. The predictor variables should be extended beyond words and sentences, and even beyond the text itself, as may be necessary to reflect all contextual variables determining comprehension.
3. The use of readability formulas to regulate or guide the production of text should be discontinued.
4. The Navy should evaluate means of changing the management of text production to ensure more usable manuals.

## REFERENCES

- Aiken, E. G. The mental quality of military personnel: Facts, theories, and remedies. Paper presented at the 16th Annual Symposium on Applied Military Psychology. The Netherlands, May 1980.
- Biersner, R. J. Reading grade levels of Navy rate training manuals and nonresident career courses (CNET Rep. 2-75). Pensacola, FL: Chief of Naval Education and Training, May 1975.
- Booher, H. Job performance aids: Research and technology state-of-the-art (NPRDC Tech. Rep. 78-28). San Diego: Navy Personnel Research and Development Center, July 1978. (AD-A057 562)
- Bormouth, J. R. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1967, 10, 291-299.
- Bormouth, J. R. Development of readability analysis. (University of Chicago Final Rep. Project No. 7.0052, Contract No. OEC-3-7070052-0326). Washington, DC: U.S. Office of Education, 1969.
- Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. Methodologies for determining reading requirements of military occupational specialties (HumRRO Tech. Rep. 73-5). Presidio of Monterey, CA: Human Resources Organization, March 1973.
- Chief of Naval Operations. Military manpower vs. hardware procurement (HARDMAN Report). Washington, DC, 1977.
- Coleman, E. B. Improving comprehensibility by shortening sentences. Journal of Applied Psychology, 1962, 46, 131-134.
- Curran, T. E. Survey of technical manual readability and comprehensibility (Tech. Rep. 77-37). San Diego: Navy Personnel Research and Development Center, June 1977. (AD-A042 335)
- Curran, T. E. (Ed). Tri-service literacy and readability: Workshop proceedings (NPRDC SR 80-12). San Diego: Navy Personnel Research and Development Center, March 1980.
- Dale, E., & Chall, J. S. A formula for predicting readability. Education Research Bulletin, 1948, 11, 37-54.
- Department of Defense. Standard for comprehensible writing. Washington, DC: Author, 1978.
- Department of Defense. Military standard for technical manuals (DOD-STD-1685 (SH)). Washington, DC: Author, 10 October 1978.
- Department of the Air Force. Publication management writing understandable publications (HQ Operating Instruction 5.2). Washington, DC: Headquarters, U.S. Air Force, March 1977.
- Department of the Army. Improving the readability of Army publications (DA Circular 310-9). Washington, DC: Author, 15 December 1978.

- Duffy, T. M. Producing technical manuals: Specifications & guidelines. Paper presented at the American Educational Research Association annual meeting. New York, March 1982.
- Duffy, T. M., & U'Ren, P. K. When does readability revision improve comprehension? (NPRDC Spec. Rep. 82-44). San Diego: Navy Personnel Research and Development Center, May, 1982 (AD-A114 935).
- Duffy, T. M., & Nugent, W. Reading skill levels in the Navy (NPRDC SR 78-19). San Diego: Navy Personnel Research and Development Center, April 1978.
- Entin, E.B., & Klare, G.R. Some inter-relationships of readability, cloze, and multiple choice scores on a reading comprehension test. Journal of Reading Behavior, 1978, 10, 417-436.
- Fass, W. & Shumacher, G. M. Effects of motivations, subject activity, and readability on the retention of prose materials. Journal of Educational Psychology, 1978, 70, 803-807.
- Feldman, M. E. The effects of learning by programmed and text format at three levels of difficulty (Unpublished dissertation). Cornell University, 1964.
- Flesch, R. F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.
- Flesch, R. F. The art of readable writing. New York: Harper, 1949.
- Gates, A. I., & MacGinitie, W. H. Gates-MacGinitie reading tests, Survey D. New York: Teachers College Press, 1965.
- General Accounting Office, United States. Improved management of maintenance manuals needed in DoD (LCD-79-105). Washington, DC: General Accounting Office, 10 July 1979.
- Goetz, E. T. Sentences in lists and in connected discourse (Tech. Rep. 3). Urbana-Champaign, IL: Laboratory for Cognitive Studies in Education, November 1975.
- Hiller, J. H. Learning from prose text: Effects of readability level, inserted question difficulty, and individual differences. Journal of Educational Psychology, 1974, 66, 202-211.
- Hooke, L., De Leo, J., & Slaughter, S. Readability of Air Force publications: A criterion referenced evaluation (AFHRL-TR-79-21). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, September 1979.
- Kern, R. P. Usefulness of readability formulae for achieving Army readability objectives: Research and state-of-the-art applied to the Army's problem. Fort Benjamin Harrison, IN: Technical Advisory Service, Army Research Institute, 13 June 1979.
- Kincaid, J. P., & Delionbach, L. Validation of the automated readability index: A follow-up. Human Factors, 1973, 15, 17-20.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel (Research Branch Report 8-75). Millington, TN: Naval Air Station, Memphis, February 1975.

- Klare, G. R. The measurement of readability. Ames, IA: Iowa State University Press, 1963.
- Klare, G. R. A manual for readable writing. Glen Burnie, MD: REM Company, 1975.
- Klare, G. R. A second look at the validity of readability formulas. Journal of Reading Behavior, 1976, 8, 129-152.
- Klare, G. R. Readability standards for Army-wide publications (Evaluation Report 79-1). Fort Benjamin Harrison, IN: US Army Administrative Center, 1979.
- Klare, G. R., Mabry, J., & Gustafson, L. The relationship of style difficulty to immediate retention and to acceptability of technical material. Journal of Educational Psychology, 1955, 46, 287-295.
- Klare, G. R., & Smart, K. Analysis of the readability level of selected USAFI instructional materials. Journal of Educational Research, 1973, 67, 176.
- Kniffen, J. D., Stevenson, C. R., Klare, G. R., Entin, E., Slaughter, S., & Hooke, L. Operational consequences of a literacy gap (AFHRL-TR-79-22). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, November 1979.
- Mackovak, W. G. Literacy skills and requirements in Air Force career ladders (AFHRL-TR-74-90). Lowry Air Force Base, CO: Air Force Human Resources Laboratory, December 1974.
- McDonald-Ross, M., & Waller, R. The Transformer. The Penrose Annual. 1976, 69, 141-152.
- Muller, W. Review of Navy research on the usability of technical manuals. In T.G. Sticht and D. Zapf (Eds.). Reading and readability research in the Armed Services (HumRRO FR-WD-CA-76-4). Alexandria, VA: Human Resources Research Organization, September 1976.
- Nelson, M. J., & Denny, E. C. The Nelson-Denny Reading Test. Revised by J. I. Brown. Boston, MA: Houghton Mifflin Co., 1960.
- Nolte, K. F. Simplification on vocabulary and comprehension in reading. Elementary English Review, 1937, 14, 119-124.
- Pearson, P. D. The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. Reading Research Quarterly, 1974-1975, 10, 155-192.
- Pressman, R. Legislative and regulatory progress on the readability of insurance policies. Washington, DC: American Institutes for Research. December 1979.
- Rankin, E. F., & Culhane, J. W. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13, 193-198.
- Reder, L. Comprehension and retention of prose: A literacy review (Tech. Rep. No. 108). Champaign, IL: Center for the Study of Reading, November 1978.
- Redish, J. Readability. In D. A. McDonald (Ed). Drafting documents in plain language. New York: Practicing Law Institute, 1979.

Smith, E. A., & Kincaid, J. P. Derivation and validation of the automated readability index for use with technical materials. Human Factors, 1960, 12, 457-464.

Sticht, T., Fox, L., Hawke, R., & Zapf, D. The role of reading in the Navy (NPRDC TR 77-40). San Diego: Navy Personnel Research and Development Center, September 1977. (AD-A044 227)

Sulit, R. A., & Fuller, J. J. Navy technical manual system (NTMS) program summary (TM-186-76-1). Bethesda, MD: Naval Ship Research and Development Center, March 1976.

Taylor, W. L. Cloze procedure: A new tool for measuring readability. Journal Quarterly, 1953, 30, 415-433.

Tuinman, J. J. & Brady, M. E. How does vocabulary account for variance on reading comprehension tests? A preliminary to an instructional analysis. Presented at the National Reading Conference, Houston, Texas, December 1973.

## **DISTRIBUTION LIST**

Chief of Naval Material (NMAT 00)  
Deputy Chief of Naval Material (Technology)  
Chief of Naval Research (Code 200), (Code 440) (3), (Code 442), (Code 448)  
Chief of Information (OI-213)  
Chief of Naval Education and Training (02), (N-2), (N-5)  
Chief of Naval Technical Training (016)  
Commander in Chief U.S. Atlantic Fleet  
Commander in Chief U.S. Pacific Fleet  
Commander Anti-Submarine Warfare Wing, U.S. Pacific Fleet  
Commander Fleet Training Group, Pearl Harbor  
Commander Naval Air Force, U.S. Atlantic Fleet  
Commander Naval Air Force, U.S. Pacific Fleet  
Commander Naval Surface Force, U.S. Atlantic Fleet  
Commander Naval Surface Force, U.S. Pacific Fleet  
Commander Naval Weapons Center  
Commander David W. Taylor Naval Ship Research and Development Center  
Commander Navy Recruiting Command  
Commander Training Command, U.S. Atlantic Fleet  
Commander Training Command, U.S. Pacific Fleet  
Commanding Officer, Fleet Anti-Submarine Warfare Training Center, Pacific  
Commanding Officer, Fleet Combat Training Center, Atlantic  
Commanding Officer, Fleet Combat Training Center, Pacific  
Commanding Officer, Fleet Training Center, San Diego  
Commanding Officer, Naval Coastal Systems Center  
Commanding Officer, Naval Underwater Systems Center  
Commanding Officer, Naval Damage Control Training Center  
Commanding Officer, Naval Education and Training Program Development Center (Technical Library) (2)  
Commanding Officer, Naval Education and Training Support Center, Pacific  
Commanding Officer, Naval Technical Training Center, Corry Station (Code 101B)  
Commanding Officer, Naval Training Equipment Center (Technical Library)  
Commanding Officer, Office of Naval Research Branch Office, Chicago (Coordinator for Psychological Sciences)  
Commanding Officer, Recruit Training Command (Academic Training Division)  
Director, Career Information and Counseling School (Code 3W34)  
Director, Defense Activity for Non-Traditional Education Support  
Director, Management Information and Instructional Activity Branch Office, Memphis  
Director, Naval Education and Training Program Development Center Detachment, Great Lakes  
Director, Naval Education and Training Program Development Center Detachment, Memphis  
Director, Training Analysis and Evaluation Group (TAEG)  
President, Naval War College (Code E114)  
Superintendent, Naval Postgraduate School  
Secretary Treasurer, U.S. Naval Institute  
Commander, Army Research Institute for the Behavioral and Social Sciences, Alexandria (PERI-ASL)  
Director, Systems Research Laboratory, Army Research Institute for the Behavioral and Social Sciences, Alexandria (PERI-SZ)  
Director, U.S. Army TRADOC Systems Analysis Activity, White Sands Missile Range (Library)  
Chief, Army Research Institute Field Unit--USAREUR (Library)  
Chief, Army Research Institute Field Unit, Fort Harrison



Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Scientific and Technical Information Office)  
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base (Technical Training Branch)  
Commander, Air Force Human Resources Laboratory, Williams Air Force Base (AFHRL/OT)  
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base (AFHRL/LR)  
Commander, 314 Combat Support Group, Little Rock Air Force Base (Career Progression Section)  
Commandant Coast Guard Headquarters  
Commanding Officer, U.S. Coast Guard Institute  
Commanding Officer, U.S. Coast Guard Research and Development Center, Avery Point  
Commanding Officer, U.S. Coast Guard Training Center, Alameda  
Defense Technical Information Center (DDA) (12)

END

FILMED

2-83

DTIC